Repeatability, Reproducibility, and Accuracy of a Novel Pushbroom Hyperspectral System

Meritxell Vilaseca,[1]* Barbara Schael,[1] Xana Delpueyo,[1] Elisabet Chorro,[2] Esther Perales,[2] Tapani Hirvonen,[3] Jaume Pujol[1]

[1]Centre for Sensors, Instruments and Systems Development (CD6), Universitat Politècnica de Catalunya, Terrassa, Barcelona, Spain

[2]Department of Optics, Pharmacology and Anatomy, Universidad de Alicante, San Vicente del Raspeig, Alicante, Spain

[3]Institute of Photonics, University of Eastern Finland, Joensuu, Finland

Received 26 April 2013; revised 28 June 2013; accepted 8 August 2013

This study analyzes the repeatability, reproducibility and accuracy of a new hyperspectral system based on a pushbroom sensor as a means of measuring spectral features and color of materials and objects. The hyperspectral system consisted of a CCD camera, a spectrograph and an objective lens. An additional linear moving system allowed the mechanical scanning of the complete scene. A uniform overhead luminaire with daylight configuration was used to irradiate the scene using d:45 geometry. We followed the guidelines of the ASTM E2214-08 Standard Practice for Specifying and Verifying the Performance of Color-Measuring Instruments *that define the standards and latest multidimensional procedures. The results obtained are analyzed in-depth and compared to those recently reported by other authors for spectrophotometers and multispectral systems. It can be concluded that hyperspectral systems are reliable and can be used in the industry to perform spectral and color readings with a high spatial resolution.* © 2013 Wiley Periodicals, Inc. Col Res Appl, 00, 000–000, 2013; Published Online 00 Month 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/col.21851

Key words: hyperspectral systems; repeatability; reproducibility; accuracy

*Correspondence to: Meritxell Vilaseca (e-mail: mvilasec@oo.upc.edu)
Contract grant sponsor: Spanish Ministry of Education and Science; contract grant numbers: DPI2008-06455-C02-01, DPI DPI2011-30090-C02-01; Contract grant sponsor: European Union.

INTRODUCTION

Hyperspectral cameras that measure the complete spectrum for each pixel of an image have appeared on the market in recent years.[1] These systems, which combine the strength of conventional imaging with that of spectroscopy, provide data with a spectral resolution larger than that obtained by means of multispectral systems. While no formal definition differentiates between hyperspectral and multispectral, the difference is usually based on the number of bands. Multispectral systems commonly use from 4 to 10 fairly narrow discrete bands with different spectral features, which is what distinguishes multispectral in the visible from conventional RGB color imaging. Some reports have suggested that less than 10 channels are usually needed[2] due to the relatively smooth spectral properties of most surfaces.[3] However, the use of hyperspectral systems using tens or even hundreds of contiguous bands is recommended when high spectral accuracy is required for a particular application or the analyzed data include narrow spectral peaks. Indeed, the use of hyperspectral systems provides significant advantages in the fields of colorimetry and spectrometry, mainly in the precise spectral characterization of nonuniform materials with complex spatial patterns.

Hyperspectral images can be obtained by means of two different kinds of devices: firstly, those commonly known as pushbroom sensors, which consist of an optical system projecting an image onto a linear array of sensors.[4] Pushbroom sensors' main components are a digital camera, a spectrograph based on a diffraction grating and an

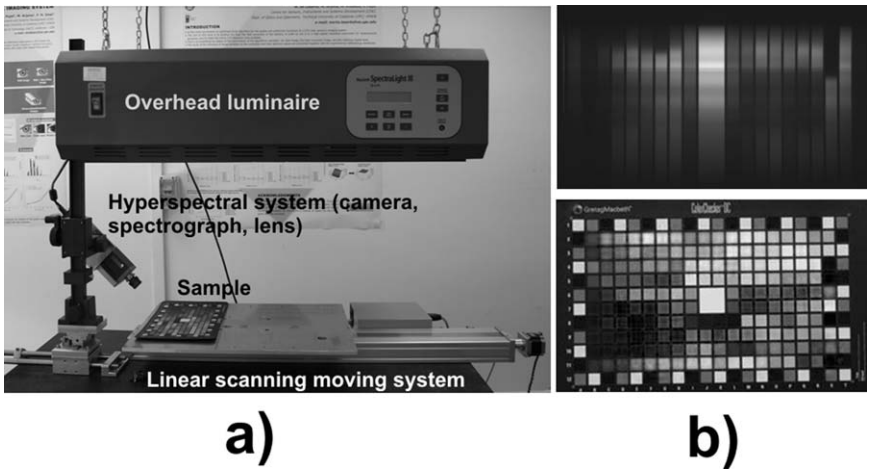

FIG. 1. (a) Experimental setup of the hyperspectral system; (b) Spectral image corresponding to a unique line of the scene (top) and spectral image of the whole scanned scene (bottom).

objective lens. In summary, a pushbroom sensor works as follows: a light source irradiates a surface or scene and the reflected light is captured by the objective lens. Next, a linear section of this radiation is selected using a slit. Finally, the spectrograph disperses the radiation, which ultimately reaches the sensor matrix. Regardless of the sensor's orientation, one can think that the scattered radiation reaches the sensor vertically, whereas the linearly sectioned radiation arrives horizontally, so that the sensor array contains spectral information in the vertical component (spectral signature) and spatial information in the horizontal plane. An additional moving system enables the mechanical scanning of the complete scene, although occasionally the scan can be performed optically. The spectral bandwidth linked to this kind of systems is usually similar or under 1 nm. On the other hand, there are also systems based on a digital camera, a liquid crystal tunable filter and an objective lens.[5,6] The spectral transmittance can be controlled electronically and changed sequentially with the use of tunable filters, thus generating monochrome images for each available spectral band, which are often captured using a 10 or 20 nm spectral step. Any of the two abovementioned systems sample the scene spectrally and spatially, creating a three dimensional (3D) cube of data (x, y, λ) with spectral information pixel by pixel, i.e., a hyperspectral image.

The aim of this study is to investigate the repeatability, reproducibility and accuracy of a hyperspectral system based on a recently commercially available pushbroom sensor as a means of measuring spectral features and color of materials and objects. Previously, spectral and color-measuring instruments' performance was exclusively investigated with the analysis of the traditional metrics based on color difference.[7–12] However, since color is a multidimensional property of a material, repeatability, reproducibility, and accuracy should be reported in terms of multidimensional parameters and not only using color differences. Another frequent problem when using color differences is that they do not follow a normal distribution, but a curve related to the chi-squared of F statistical distributions.[13] Consequently, in the analysis by means of color differences the mean, mode, and median of typical color difference data will differ because of lack of normality. Indeed, the standard terms and latest multidimensional procedures useful for the purpose of this study were not properly defined and specified until the publication of the ASTM E2214-08 *Standard Practice for Specifying and Verifying the Performance of Color-Measuring Instruments*.[14] In addition, some authors have recently published useful information to implement the mathematical tools required to compute all these new recommended metrics after using these procedures to characterize some instruments, from conventional and multiangle spectrophotometers[15–17] to multispectral systems.[18]

This study follows the guidelines specified in the ASTM E2214-08 standard to set up a novel hyperspectral system. This article is structured as follows: firstly, the experimental setup of the system used is characterized; the methodology to analyze repeatability, reproducibility and accuracy is described next. The results section reports the repeatability, reproducibility and accuracy of the hyperspectral system. The discussion includes the comparison of the results regarding standard spectrophotometers and multispectral systems with those published by other authors. Finally, the most relevant conclusions are listed.

EXPERIMENTAL SETUP

The hyperspectral system is shown in Fig. 1(a). A uniform and calibrated light source SpectraLight III overhead luminaire with daylight D65 configuration (T_c = 6458 K) designed for critical color evaluation irradiated the scene using d:45 geometry (Diffuse illumination/45° observation angle). The viewing direction was kept constant from one end to another of the scan line. The reflected light was later captured with the hyperspectral system, which consisted of a 14-bit digital CCD camera (AVT Pike F-210B), a spectrograph (ImSpector V10E), and an objective lens (Cinegon 1.8/16). Thanks to the dispersion obtained with the diffraction grating of the spectrograph,

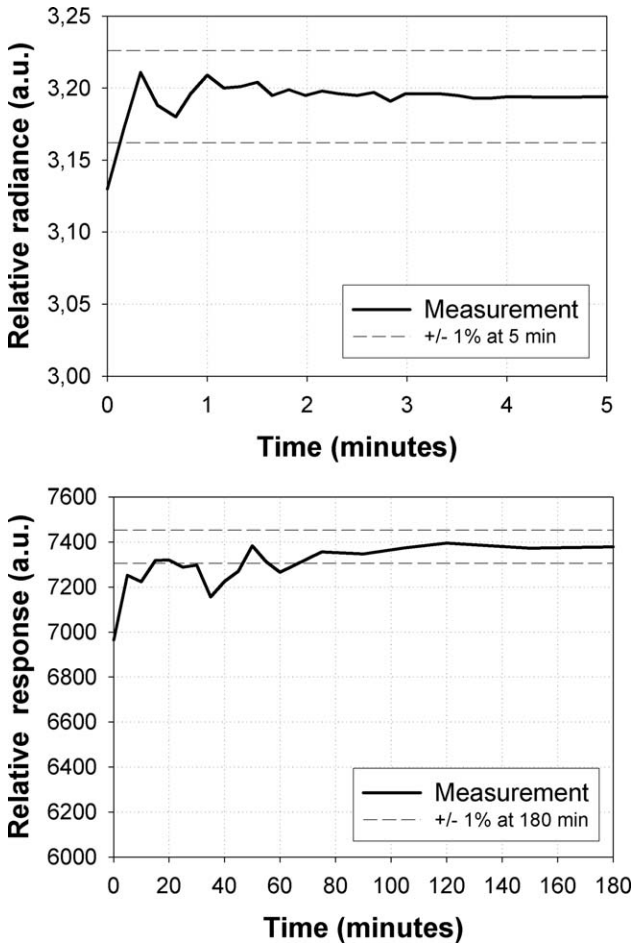

FIG. 2. (a) Relative radiance and (b) relative camera response measured over time. Reference levels of ±1% regarding the last measurement are also plotted.

the system provided spectral information in terms of digital levels of every pixel corresponding to a unique line of the scene. Next, an additional linear moving system enabled the mechanical scanning of the complete scene [Fig. 1(b)].

Some preliminary tests were carried out in order to ensure the good performance of the system. The light source stability was checked by repeating measurements of its radiance over time. A variation smaller than ±1% was achieved after some seconds it was switch on [Fig. 2(a)]. However, it must be taken into account that although the lamp stability was reached very soon, the stability linked to the whole hyperspectral system was not, probably due to the increase in temperature that takes place at the camera sensor. The drift in the response of the hyperspectral system over time can be seen in Fig. 2(b). In this case, the time needed to obtain a variation smaller than ±1% was of 70 min. However, about 90 min were necessary to obtain an almost perfect stabilization, and consequently, this time lapse was always considered before starting measurements.

On the other hand, to avoid light coming from other secondary sources, the background on which the sample was placed and the surrounding area were covered with a neutral matte fabric. Furthermore, although the luminaire included a diffuser located below the bulbs and was installed no closer than 2 feet to the nearest walls to minimize light level uniformity problems as the manufacturer suggests, the irradiance profile corresponding to the measurement area was not perfectly uniform and showed a variation of approximately ±3% (Fig. 3). Specifically, reflectance factors for each pixel (i) of the captured line were computed taking into account the image of a calibrated white plate (Gigahertz-Optik's BN-R98-SQ12, 254 × 254 mm) and a dark current image as follows:

$$r(i,\lambda)=\frac{DL(i,\lambda)-DL_{Dark}(i,\lambda)}{DL_{White}(i,\lambda)-DL_{Dark}(i,\lambda)}\cdot cal_{White}(\lambda) \qquad (1)$$

where r is the reflectance factor, DL is the digital level of the image being corrected, DL_{Dark} is the digital level of the dark current image, DL_{White} is the digital level of the image of the calibrated white plate and cal_{White} is the calibrated reflectance of the white plate provided by the manufacturer. It must be remembered that light reaching each pixel of the captured line is dispersed so that finally information on every wavelength (λ) is available.

The hyperspectral camera allowed measurements from 400 nm to 1000 nm approximately with a 0.7 nm step, although a bandwidth of 4 nm was finally selected to obtain an acceptable signal to noise ratio.

METHOD

Repeatability is defined in the ASTM E2214-08 as the most important specification in a color-measuring instrument since colorimetry is primarily a relative or differential measurement. Repeatability tests explain how well an instrument executes the readings of the same specimen by using a single operator over a period of time. It is usually assumed that the standard deviation is a good estimate of repeatability. The ASTM E2214-08 standard defines reproducibility as the second most important specification in a color-measuring instrument. Reproducibility is a form of repeatability in which one or more measurement parameters have been systematically changed, i.e., the sample is different, the procedures or instrument are different, or the time frame is prolonged. In this study we will use inter-

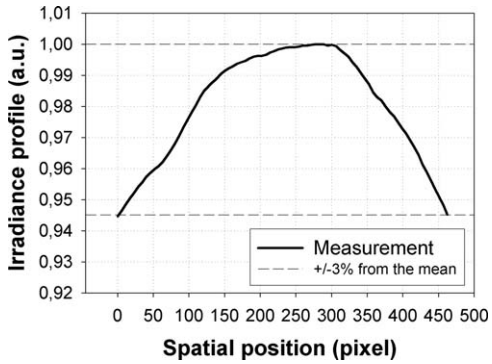

FIG. 3. Spatial uniformity at the sample level. The variability of the irradiance profile is smaller than ±3%.

model reproducibility, which describes the reproducibility between two or more instruments of differing design. Finally, the same standard defines accuracy as generally the least significant parameter to characterize the performance of a color-measuring instrument. Accuracy is often defined as the conformance of a series of readings to the true value. In a spectroradiometer, two scales can be assigned nominal values and tested against standard values: the radiometric and the wavelength scale.

Analysis of Repeatability

Based on the former definitions and on the guidelines established in the ASTM E2214-08, we performed measurements on a calibrated white plate (Gigahertz-Optik's BN-R98-SQ12, 254 × 254 mm made of a highly durable, light and temperature stable PTFE material with an optimal diffuse reflectance [Lambertian reflector] and a spectrally neutral reflectance of 98 ± 1% from 400 nm to 800 nm) and used univariate and multivariate metrics to analyze the repeatability of the hyperspectral system. Fifty consecutive readings were taken to account for short-term repeatability, 50 in two consecutive days for medium-term repeatability, and 50 during 5 weeks for long-term repeatability. Multivariate methods used included the study of the following parameters: $\Delta R_{\lambda,2\sigma}$ (at 440, 560 and 650 nm), $\Delta X_{2\sigma}$, $\Delta Y_{2\sigma}$, $\Delta Z_{2\sigma}$, $\Delta L^*_{2\sigma}$, $\Delta a^*_{2\sigma}$, $\Delta b^*_{2\sigma}$, and $\Delta E^*_{2\sigma}$, all of them representing twice the standard deviation with regard to reflectance at a specified wavelength, the tristimulus values or the CIELAB color coordinates and color difference. In addition, we used $\sum \sigma_W^2$, calculated as the sum of the variances of the tristimulus values ($\sigma_X^2+\sigma_Y^2+\sigma_Z^2$), and V_{XYZ} and V_{Lab}, computed as the volume of the covariance matrix which describes an ellipsoid in a 3D distribution of color difference data. These volumes are related to the distribution of color difference data in the XYZ and LAB spaces, respectively. Therefore, a smaller volume correlates with lower variability in the results, i.e., better repeatability. On the other hand, the univariate parameters used to account for repeatability were traditional metrics based on color difference taken from the mean value (ΔE^*_{ab}, ΔE^*_{94} and ΔE^*_{00}), the root mean square error (RMSE) and the weighted RMSE (wRMSE), in which the spectral reflectance is weighted using a weighting function w_λ defined as the sum of standard observer functions ($x^J + y^J + z^J$) (see Refs. 14 and 15 for a more detailed information of all these metrics). The drift shown by the instruments analyzed must be also considered in the analysis of repeatability. This drift can be investigated by studying the 50 consecutive readings obtained in the analysis of short-term repeatability for some parameters (for instance L^*, a^*, and b^*).

Analysis of Reproducibility

To account for inter-model reproducibility, we compared the readings obtained with the hyperspectral system with those measured using a conventional telespectroradiometer (Photo Research PR-655 SpectraScan® Spectroradiometer), which provides results from 380 nm to 780 nm. The spectral accuracy of the telespectroradiometer PR-655 is of ± 1 nm while its luminance accuracy tested against a NIST luminance standard is of ± 2%. We used two different sets of samples: 12 glossy BCRA ceramic tiles CCS-II (Ceram Color Standards Series II) and 24 matte patches (X-Rite Munsell CCRC color checker chart). To obtain values comparable to those obtained by means of the repeatability analysis, 50 consecutive readings for each instrument were taken. Next, a multivariate statistical analysis using the Hotelling and the inter-comparison tests was performed. The Hotelling's T^2 metric is based on the computation of the 3D volume defined by the square roots of the eigenvalues of the covariance matrix of color coordinate difference values (ΔL^*, Δa^*, Δb^*). The calculated metric is therefore a description of the acceptance or tolerance volume of an instrument in terms of color coordinates for a given statistical significance. In this study, comparisons were considered significant with P values under 0.05 (95% confidence interval). The inter-comparison test, derived from propagation of errors and the chi-squared statistical distribution, computes interval estimates for the component differences ΔL^*, Δa^*, Δb^*, ΔC^*, ΔH^*, and ΔE^*; ΔE^* is usually of most interest for making industrial color measurements. In this case, if the total color difference average (ΔE^*) is higher than the critical value ($t_{\Delta E}$), which is the statistical benchmark for this test, the difference is considered significant. For the Hotelling and inter-comparison tests, significant differences imply that differences between values measured by one instrument compared to those obtained with the other are due to systematic errors, but not exclusively to random errors (see Refs. 14, 16, 17, and 18 for more detailed information on these tests). To compare reproducibility and repeatability results some of the univariate metrics such as color differences (ΔE^*_{ab} and ΔE_{00}) and the RMSE were also used.

The reflectance factors from 400 nm to 700 nm ($\Delta\lambda$ = 10 nm), illuminant CIE D65 and CIE 10° observer were used to compute the color data in all cases.[14–16]

Analysis of Accuracy

With regard to accuracy, the ASTM E2214-08 recommends the wavelength and radiometric scales, both of which can use standard values for comparison. The wavelength scale includes determining the centroid wavelength and spectral bandwidth that correspond to several peaks. The centroid wavelength corresponds to the maximum value of the fitted curve, whereas the spectral bandwidth is the full width at half maximum (FWHM). An averaged observed wavelength plus/minus a tolerance, defined as twice the standard deviation, is calculated from 10 replicates. The wavelength accuracy is the difference between the true wavelength and the average observed reading. We used a mercury-argon lamp (Ocean Optics Inc. HG-1 Mercury Argon Calibration Source) with several spectral

peaks in the 253–922 nm range as a known physical standard, and the centroid wavelength and spectral bandwidth were determined by Gaussian regression.

On the other hand, the white level, black level, and linearity must be analyzed in the radiometric scale. In this study, the accuracy corresponding to the white level was tested by direct comparison with a primary standard of reflectance: the calibrated Spectralon® diffuse reflectance standard Labsphere SRS-99-020. Radiometric accuracy was next computed as the difference between the true value (calibration provided by the manufacturer) and the averaged observed value from 10 readings at the following wavelengths: 450, 550, and 650 nm. The standard deviation of the observed values was also computed and the expanded uncertainty linked to radiometric accuracy was then calculated as the uncertainty of the reflectance standard (which is of 0.005) and that of the hyperspectral system combined in quadrature at the 95% confidence level, i.e., with a coverage factor of 2.[19] The accuracy of the black level was tested by measuring the reflectance by means of the hyperspectral system with the cap of the objective lens on. Ten replicates were also performed. These measurements are performed to demonstrate that the optical zero of the system is less than a specified value, usually 0.0005. Linearity was finally tested using the matte neutral patches of the X-Rite Munsell CCRC color checker chart. Specifically, there are 6 patches: white, light gray, light-medium gray, medium gray, dark gray, and black. A linear regression was carried out between calibrated and observed values in a photometric scale, considering each line segment between every two pairs of samples independently. Following ASTM E2214-08 recommendations, the slopes must be compared with the expected value of 1.0. The luminous reflectance factors (Y) of the 6 former patches measured with the hyperspectral system were computed using the CIE C illuminant and CIE-2° observer, since these were the values provided by the manufacturer for the patches analyzed. Ten consecutive readings of the spectral reflectance of each patch were also considered for this purpose. The maximum absolute difference plus/minus combined uncertainty in the slope due to uncertainty in the calibrated and observed values was also reported; it was calculated using the summation in quadrature (square root of the sum of the squares) of all absolute differences of the slopes of the line segments.

RESULTS

Analysis of Repeatability

Repeatability data are shown in Table I for the hyperspectral system. A first analysis of the results suggests that the instrument provides good levels of repeatability in most metrics used. For users requiring spectral data, it is of interest to highlight that the largest variability in reflectance values is obtained at longer wavelengths ($\Delta R_{650,2\sigma}$), probably because the sensitivity of the sensor at these wavelengths is considerably lower. In the case of users requiring colorimetric data, it can be established that color differences are always below 1. Figure 4 shows the data provided by the instrument in terms of L^*, a^*, and b^* using 50 consecutive readings obtained in the analysis of short-term repeatability, which does not appear to be associated with a drift in the results.

TABLE I. Results for short-, medium-, and long-term repeatability of the hyperspectral system using multivariate and univariate metrics.

Metric	Repeatability		
Multivariate	Short-term	Medium-term	Long-term
$\Delta R_{440,2\sigma}$	0.0017	0.0105	0.0223
$\Delta R_{560,2\sigma}$	0.0011	0.0112	0.0246
$\Delta R_{650,2\sigma}$	0.0017	0.0146	0.0345
$\Delta X_{2\sigma}$	0.0942	1.1504	2.5114
$\Delta Y_{2\sigma}$	0.1019	1.2096	2.6609
$\Delta Z_{2\sigma}$	0.1234	1.1606	2.5197
$\Delta L^*_{2\sigma}$	0.0397	0.4735	1.0549
$\Delta a^*_{2\sigma}$	0.0275	0.0572	0.2340
$\Delta b^*_{2\sigma}$	0.0498	0.1874	0.5387
$\Delta E^*_{2\sigma}$	0.0694	0.5124	1.2074
$\Sigma\sigma^2_w$	0.0086	1.0334	4.934
V_{XYZ}	4.7589E-05	4.9436E-03	1.294E-01
V_{Lab}	2.0543E-05	2.1689E-03	5.871E-02
Univariate			
ΔE^*_{ab}	0.031	0.191	0.537
ΔE^*_{94}	0.031	0.191	0.535
ΔE_{00}	0.031	0.130	0.399
RMSE (Absolute)	0.0009	0.0046	0.0127
wRMSE (Absolute)	0.0009	0.0048	0.0123
RMSE (%)	0.1367	0.6631	1.8667

Analysis of Reproducibility

As established in the preceding section, we measured reflectance factors of equivalent regions of the 24 matte patches corresponding to the CCRC color chart and the 12 glossy BCRA tiles using the hyperspectral system and the tele-spectroradiometer PR-655. Table II shows the color differences and RMSE values encountered between the two instruments when the two sets of samples are analyzed independently. Additionally, Fig. 5 shows the CIELAB partial differences (Δa^* vs. Δb^*, and ΔL^* vs. ΔC^*) between both instruments. As expected, the results suggest that reproducibility of the hyperspectral system is worse than repeatability, even when long-term results are considered. In general, darker patches and those having high reflectance factors in the yellow region of the visible spectrum reported worse reproducibility results (the latter in terms of color differences). It can be inferred from the figure that the clusters are spread around the color difference space depending on the patch analyzed. Δb^* differences are generally larger than Δa^* differences, and a concentration of data in the negative part of the Δb^* dimension exists. Since the partial differences have been computed by subtracting the PR-655 values from those of the hyperspectral camera (for instance: a^*_{Hyp}-$a^*_{PR\text{-}655}$) it can be concluded that b^*_{Hyp} values are smaller for many samples

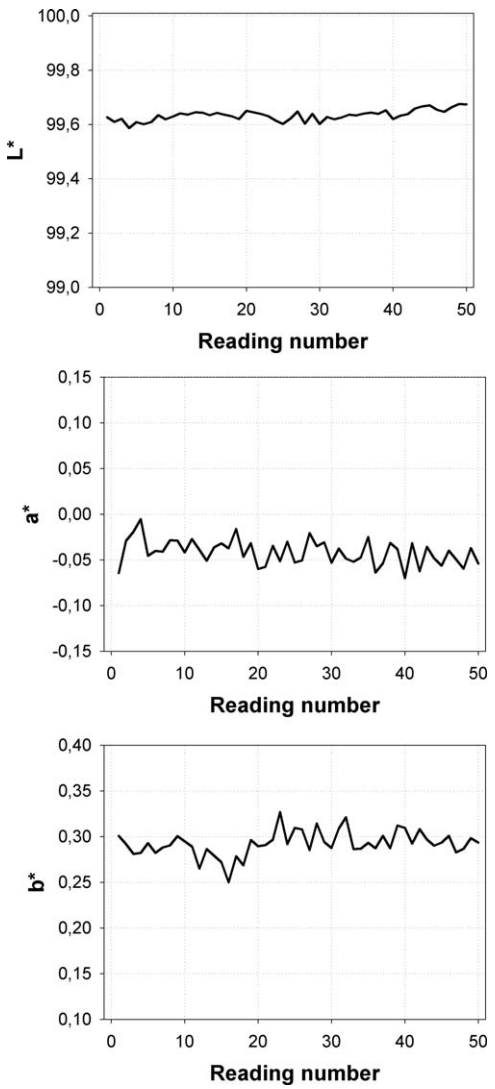

FIG. 4. L^*, a^*, and b^* vs. reading number for the hyperspectral camera. The 50 consecutive readings obtained in the short-term analysis are considered.

than those measured by the PR-655. This would partially explain the larger differences found in the yellow region of the electromagnetic spectrum. Performing the same analysis in the ΔL^* vs. ΔC^* graph, positive values are generally found for most patches in the vertical direction, meaning that L^*_{Hyp} is generally higher than $L^*_{\mathrm{PR\text{-}655.}}$

Moreover, differences were much larger for the glossy BCRA tiles than for the CCRC patches. Most CCRC samples reported quite acceptable color differences (ΔE^*_{ab} similar to 2 and ΔE_{00} to 1) and RMSE values close to 6%. However, the mean color differences for the BCRA tiles were similar to 6 (ΔE^*_{ab}) and 4 (ΔE_{00}), while the mean RMSE was 23%. Figure 6 shows the reflectance spectra of several measured samples from the two sets obtained with the hyperspectral system as well as the tele-spectroradiometer PR-655. Even if the profiles are very similar for any BCRA sample, it is clear that an offset in the curve of the hyperspectral system with respect to the PR-655 exists. The bias is always positive, and for this reason the lightness codified by the camera L^*_{Hyp} is higher than that measured by the tele-spectracolorimeter.

Finally, Table III summarizes the results of the multivariate Hotelling test. Similarly, Table IV shows the results obtained using the inter-comparison test. The results reported by the Hotelling test suggest that there are significant differences between the data measured by both systems ($P < 0.001$). The same conclusion is reached in the inter-comparison test, since ΔE^*_{ab}, which represents the average of the color differences, is larger than the critical value $t_{\Delta \mathrm{E.}}$

Analysis of Accuracy

The strongest spectral peaks of the mercury-argon lamp within the range of interest of this study, i.e., between 400 nm and 700 nm, are located at the following wavelengths: 404.656 nm, 435.833 nm, 546.074 nm, 576.960 nm, 579.066 nm, and 696.543 nm. The two lines corresponding to 576.960 nm and 579.066 nm are indistinguishable by the hyperspectral camera because they are very close to each other; consequently, we considered them together and taking into account their relative intensities, the new maximum of intensity corresponding to the centroid wavelength when performing the regression was expected to be found at 578.021 nm. Table V summarizes the results of wavelength accuracy for the hyperspectral camera. The results obtained prove that the hyperspectral camera has good accuracy in terms of the wavelength with an agreement better than 1 nm at almost all peaks.

Table VI shows the results obtained in the radiometric scale at the white level. Even if radiometric accuracy for the white level is very good in all cases, it is higher and therefore worse at larger wavelengths. Table VII shows the results obtained at the black level. All wavelengths had very low reflectance values, below the 0.0005 established in the ASTM E2214-08 standard. The linear regressions between calibrated and observed values in terms of Y, considering each line segment between every two pairs of samples independently are shown in Fig. 7. The slopes of each line segment (labeled from 1 to 5) are

TABLE II. Color differences (ΔE^*_{ab}, ΔE_{00}) and RMSE values (in absolute and percentage terms) obtained between the hypespectral camera and the PR-655 tele-spectroradiometer when the 24 matte patches corresponding to the CCRC color chart and the 12 glossy BCRA tiles are measured.

Samples	ΔE^*_{ab}	ΔE_{00}	RMSE (Absolute)	RMSE (%)
24 matte CCRC patches				
Mean	2.252	1.181	0.0082	6.5422
Standard deviation	1.641	0.680	0.0035	4.5494
Maximum	5.361	3.134	0.0142	14.9625
Minimum	0.284	0.179	0.0015	0.7844
12 glossy BCRA tiles				
Mean	6.428	4.339	0.0432	23.0598
Standard deviation	2.835	1.795	0.0075	10.7745
Maximum	10.683	6.891	0.0562	41.2766
Minimum	1.855	1.124	0.0293	4.9986

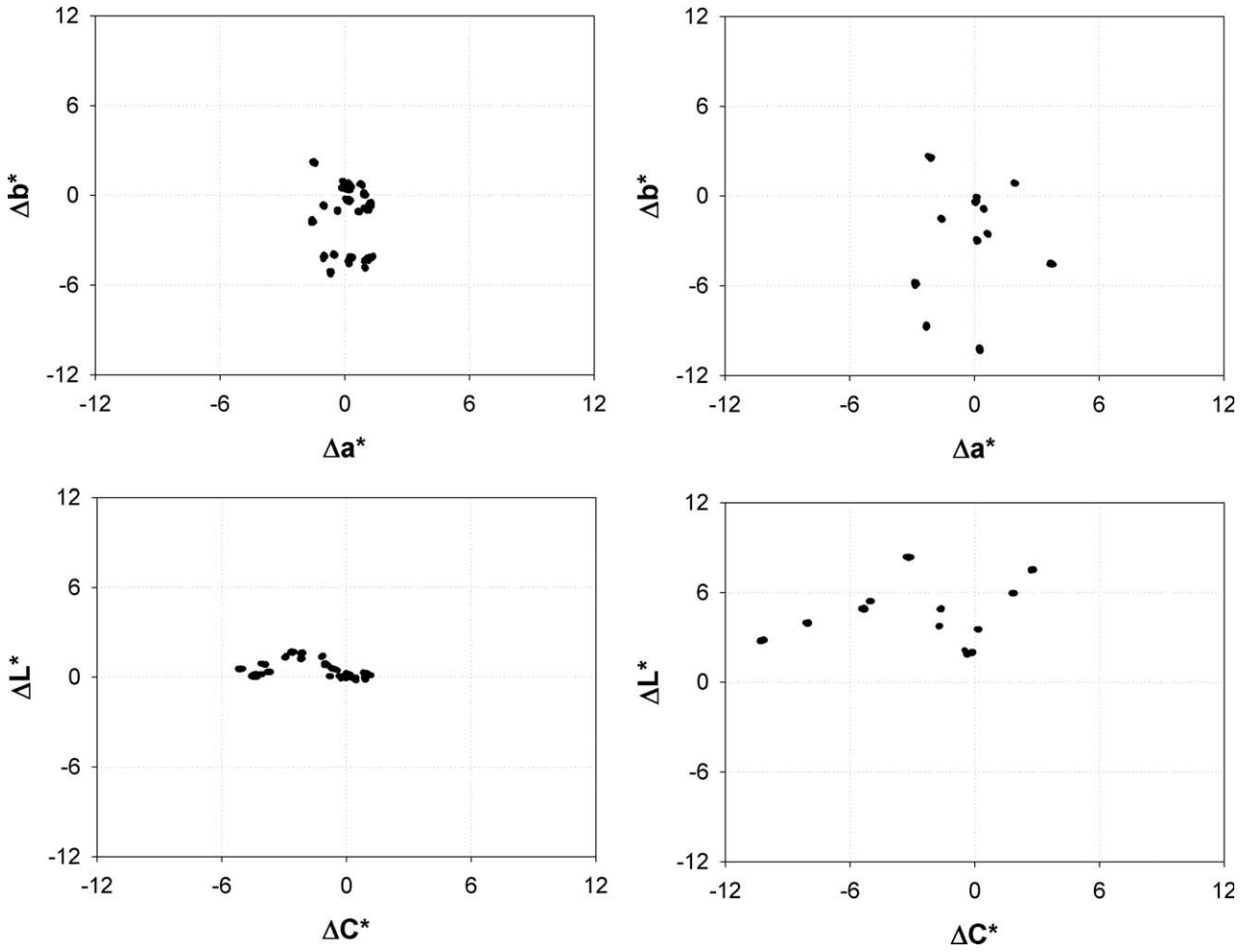

FIG. 5. Partial color differences (Δa^* vs. Δb^* and ΔL^* vs. ΔC^*) found between the hyperspectral camera and the PR-655 tele-spectroradiometer for (a) the 24 matte patches corresponding to the CCRC color chart and (b) the 12 glossy BCRA tiles.

reported in Table VIII; all are very close to 1. However, the line segment number 5 has the maximum absolute difference. This behavior was already expected because it corresponds to the darker pair of samples analyzed, i.e., the black and dark gray patches; therefore, the camera is more affected by noise. On the other hand, the maximum absolute difference plus/minus combined uncertainty in the slope due to uncertainty in the calibrated and observed values was 0.0734 ± 0.0116. The maximum absolute difference corresponds to the same line segment.

DISCUSSION

Wyble and Rich's[15] study of repeatability compared the repeatability of 12 commercial spectrophotometers that included eight integrating sphere instruments and four hand-held bidirectional devices. Based on the values reported by these authors, we can state that the hyperspectral camera has a precision similar to the majority of instruments used for measuring color.

With regard to the reproducibility analysis using the glossy BCRA ceramic tiles, larger differences were found between the hyperspectral camera and the PR-655 tele-spectroradiometer than those reported with the CCRC chart patches. The geometry measurement (D/45) and the surface of both samples might explain these results: whereas the BCRA tiles are glossy, those of the CCRC color chart are matte. If the observation is performed at 45° the positioning of the sample with respect to the light source as well as the instrument is more critical and the gloss might contribute to a higher variability among the results. Furthermore, these instruments do not have identical configurations, which supports the former explanation and may explain the larger differences found for the BCRA tiles (Fig. 8). The hyperspectral camera used allows the measurement of the complete spectrum for each pixel of an image, which is achieved by measuring only one line on the scene at once (the other direction is used to account for spectral information). Next, an additional mechanical moving system allows the linear scanning of the part of the sample to be analyzed. In contrast, the PR-655 does not need the scanning system, since it already has a viewing field of 1°. Therefore, energy coming from this region is integrated and assumed as uniform. Both instruments measured the same region of the

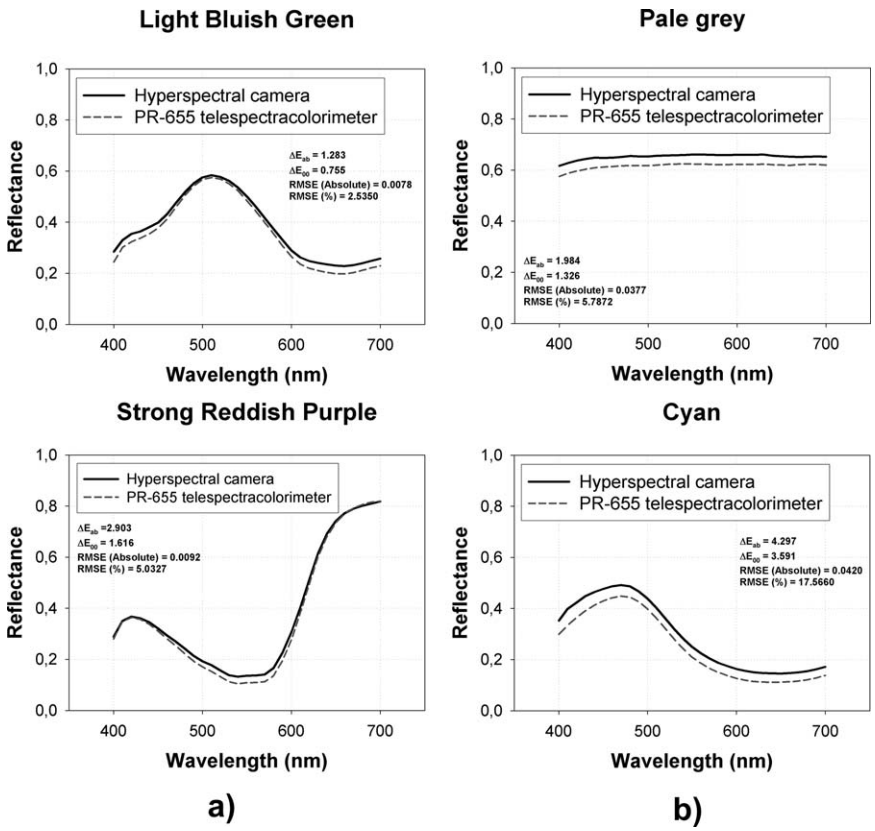

FIG. 6. Reflectance spectra of several samples measured with the hyperspectral camera and the PR-655 tele-spectroradiometer corresponding to (a) the 24 matte patches corresponding to the CCRC color chart and (b) the 12 glossy BCRA tiles.

samples for the whole study; the gloss probably affected both instruments differently and for this reason worse results were obtained with the BCRA. Non-uniformities of the illumination on the sample measured could also play an important role in the results. Even if the overhead luminaire is designed to be perfectly uniform and incorporates a diffuser placed below the light bulbs, it does not provide a perfect uniform field of illumination (a variability of $\pm 3\%$ is measured along the irradiance profile). Meanwhile the hyperspectral system applies a flat-field algorithm to correct this pixel by pixel, the PR-655 tele-spectroradiometer does not, and a perfect uniform field of illumination is assumed in this last case to account for the reflectance factor. We should point out that although the multidimensional procedures proposed by the ASTM E2214-08 have strongly been recommended to perform reproducibility analysis between instruments for color measurements, the statistical tolerance using this methodology is generally very strict as already reported by other authors, who also found differences among 10 commercial spectrophotometers[16] and 4 multiangle spectrophotometers[17] at the 5% significance level. The extraordinary repeatability of both instruments, greater than one would expect by chance variation of their readings, makes it almost impossible to obtain identical data. Consequently, the distributions of the repeatability data from both instruments form very small ellipsoids and they rarely overlap.

TABLE III. Multivariate Hotelling test performed between the hyperspectral camera and the PR-655 tele-spectroradiometer when the 24 matte patches corresponding to the CCRC color chart and the 12 glossy BCRA tiles are measured.

Hotelling test	
24 matte CCRC patches	
Sample size	1200
Degrees of freedom	3
T^2	2246.3282
P value	< 0.001
12 glossy BCRA tiles	
Sample size	600
Degrees of freedom	3
T^2	4581.5575
P value	< 0.001

TABLE IV. Multivariate intercomparison test between the hyperspectral camera and the PR-655 tele-spectroradiometer when the 24 matte patches corresponding to the CCRC color chart and the 12 glossy BCRA tiles are measured.

Inter-comparison test	
24 matte CCRC patches	
$t_{\Delta E}$	0.484
ΔE^*_{ab}	2.258
12 glossy BCRA tiles	
$t_{\Delta E}$	0.684
ΔE^*_{ab}	6.448

TABLE V. Wavelength accuracy of the hyperspectral system (SD: standard deviation).

True wavelength (nm)	Observed centroid wavelength ± 2·SD (nm)	Observed spectral bandwidth ± 2·SD (nm)	Wavelength accuracy (nm)
404.656	403.5 ± 0.9	3.5 ± 0.2	1.2
435.833	435.9 ± 0.4	3.4 ± 0.1	0.1
546.074	546.5 ± 0.5	3.8 ± 0.7	0.4
578.021	578.1 ± 0.7	5.3 ± 0.2	0.3
696.543	695.6 ± 1.5	4.0 ± 2.1	1.0

In conclusion, the variance in repeatability is insignificant compared to variances between patches and instruments.

Finally, the accuracy of the results reported in this study was affected by the gloss of the BCRA ceramic tiles. Therefore, even though at the beginning and following the guidelines of the ASTM E2214-08 we expected to test the linearity of the system in a photometric scale by means of the five neutral BCRA ceramic tiles, the linearity was finally tested by using the matte neutral patches of the CCRC color chart instead. A good performance was found for the hyperspectral system in the wavelength and radiometric scales. Wavelength accuracy was smaller than 1 nm for almost the entire range of tested visible wavelengths. The radiometric scale obtained very accurate readings of the white level; readings associated with longer wavelengths were larger. The accuracy of the black level was smaller than 0.0005 at all wavelengths. Finally, the hyperspectral system was very linear in a photometric scale; worse results were obtained only when very dark samples were taken into account.

The comparison between hyperspectral and multispectral systems deserves special attention. These systems have been used in recent years to obtain images with color and spectral information pixel by pixel by adding few acquisition channels or spectral bands in front of a digital camera. Chorro *et al.*[18] reported mean color differences (ΔE_{ab}) between colorimetric and multispectral systems and a reference tele-spectroradiometer (Photo Research PR-655 SpectraScan® Spectroradiometer). Values of 5.20 and 2.96 were obtained with a 3-channel (RGB channels) and a 7-channel configuration, respectively; both systems used a 12 bit cooled digital camera when measuring the patches of the CCDC chart (180 patches). The mean color differences obtained here are

TABLE VI. Radiometric accuracy corresponding to the white level of the hyperspectral system (SD: standard deviation).

Wavelength (nm)	True reflectance	Observed reflectance ± SD	Radiometric accuracy	Uncertainty (95%)
450	0.988	0.990 ± 0.001	0.002	0.005
550	0.988	0.992 ± 0.001	0.004	0.005
650	0.989	0.996 ± 0.001	0.007	0.005

TABLE VII. Radiometric accuracy corresponding to the black level of the hyperspectral system (SD: standard deviation).

Wavelength (nm)	Observed reflectance ± SD
450	0.000068 ± 0.000041
550	0.000040 ± 0.000027
650	0.000061 ± 0.000072

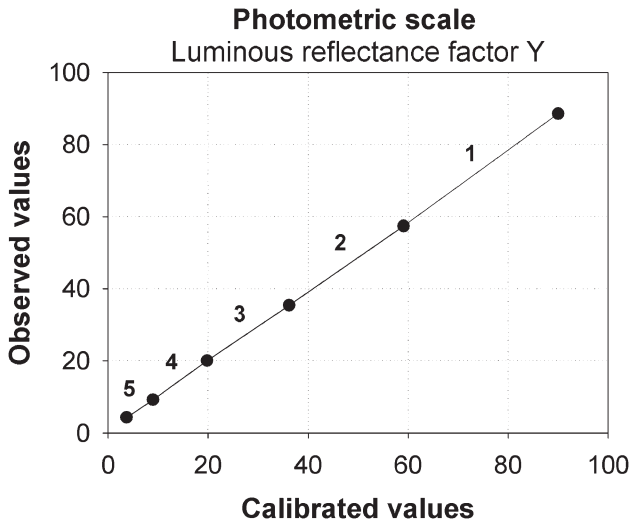

FIG. 7. Linear regression between accepted luminous reflectance factors (Y) provided by the manufacturer and those measured with the hyperspectral camera (observed values) when each line segment (labeled from 1 to 5) is considered independently.

TABLE VIII. Slopes of the linear regressions between calibrated luminous reflectance factors provided by the manufacturer and those observed with the hyperspectral system.

Line segment	Observed slope	Expected slope	Absolute difference	Difference (%)
1	1.0086	1	0.0086	0.86
2	0.9613	1	0.0387	3.87
3	0.9319	1	0.0681	6.81
4	1.0052	1	0.0052	0.52
5	0.9266	1	0.0734	7.34

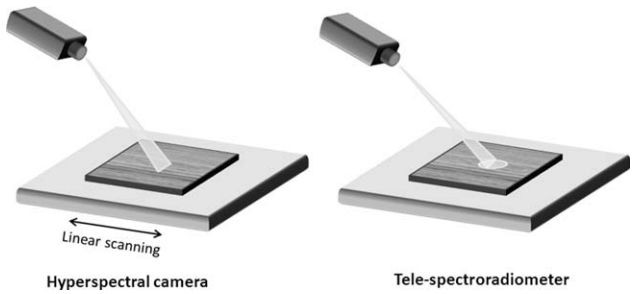

FIG. 8. Differences between the hyperspectral camera and the PR-655 tele-spectroradiometer regarding the geometric configuration.

lower if one considers the CCRC color chart (more similar to CCDC patches, which are generally matte). Indeed, hyperspectral systems using a spectrograph are more precise and accurate than multispectral systems since they are closer to a conventional spectrometric configuration with many spectral readings (every 4 nm in the case of the instrument analyzed in this study). Moreover, the hyperspectral system includes a 14 bit camera, whereas in the colorimetric and multispectral configurations a 12 bit camera was used.

CONCLUSIONS

The hyperspectral system obtained good repeatability and accuracy results and adequate reproducibility data, very similar to those already reported by other authors in commercially available spectrophotometers. Reproducibility using glossy samples was the only exception, probably attributable to the different impact of gloss depending on the configuration of the instrument used and the nonuniformities in the field of illumination.

In conclusion, hyperspectral systems are reliable and maintain and even improve reading precision and accuracy while overcoming some of the current existing limitations of conventional instruments and multispectral cameras; consequently, they can be used in the industry to perform spectral and color readings with a high spatial resolution.

1. Chang CI. Hyperspectral Data Exploitation: Theory and Applications. New York: Wiley; 2007.
2. Vhrel MJ, Trussell HJ. Filter considerations in color correction. IEEE Trans Image Process 1994;3:147–161.
3. Imai FH, Wyble DR, Berns RS, Tzeng D. A Feasibility study of spectral color reproduction. J Imaging Sci Technol 2003; 47:543–553.
4. Gupta R, Hartley RI. Linear pushbroom cameras. IEEE Trans Pattern Anal Mach Intell 1997;19:963–975.
5. Hardeberg JY, Schmitt F, Brettel H. Multispectral color image capture using a liquid crystal tunable filter. Opt Eng 2002;41:2532–2548.
6. Lopez-Alvarez MA, Hernandez-Andres J, Romero, J. Developing an optimum computer-designed multispectral system comprising a monochrome CCD camera and a liquid-crystal tunable filter. Appl Opt 2008; 47:4381–4390.
7. Billmeyer FW Jr. Comparative performance of color-measuring instruments. Appl Opt 1969;8:775–783.
8. Billmeyer FW Jr, Campbell ED, Marcus RT. Comparative performance of color-measuring instruments; Second report. Appl Opt 1974;13: 1510–1518.
9. Billmeyer FW Jr, Alessi PJ. Assessment of color-measuring instruments. Color Res Appl 1981;6:195–202.
10. Corciolani G. Vichi A. Repeatability of colour reading with a clinical and laboratory spectrophotometer. Int Dent 2006;8:62–70.
11. Rodgers J, Wolf K, Willis N, Hamilton D, Ledbetter R, Stewart C. A comparative study of color measurement instrumentation. Color Res Appl 1994;19:322–331.
12. Raggi A, Barbiroli G. Colour-difference measurement: The sensitivity of various instruments compared. Color Res Appl 1993;18:11–27.
13. Hotelling H. The Generalization of the Student's Ratio. Annals Math Stat 1931;2:360–378.
14. ASTM E2214-08. Standard practice for specifying and verifying the performance of color-measuring instruments. American Society for Testing and Materials, USA; 2008.
15. Wyble DR, Rich DC. Evaluation of methods for verifying the performance of color-measuring instruments. I. Repeatability. Color Res Appl 2007;32:166–175.
16. Wyble DR, Rich DC. Evaluation of methods for verifying the performance of color-measuring instruments. II. Interinstrument reproducibility. Color Res Appl 2007;32:176–194.
17. Perales E, Chorro E, Viqueira V, Martínez-Verdú FM. Reproducibility comparison among multiangle spectrophotometers. Color Res Appl 2013;38:160–167.
18. Chorro E, Vilaseca M, Herrera JA, Perales W, Martínez-Verdú FM, Pujol J. Evaluation of repeatability and reproducibility levels for color measurement obtained by digital imaging capture devices. Proceedings of the Fifth European Conference on Colour in Graphics, Imaging, and Vision (Joensuu, Finland). 2010;1:366–371.
19. Taylor BN, Kuyatt CE. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. National Institute of Standards and Technology, NIST Technical Note 1297, USA; 1994.